ACS Medicinal Chemistry Letters

Letter

Cite This: ACS Med. Chem. Lett. 2018, 9, 204−209

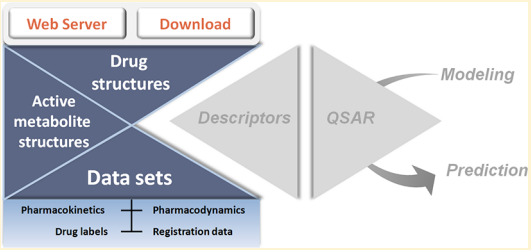
Data Sets Representative of the Structures and Experimental Properties of FDA-Approved Drugs

Dominique Douguet*

Université Côte d'Azur, Inserm, CNRS, IPMC, 660 Route des Lucioles, 06560 Valbonne, France

Supporting Information

ABSTRACT: Presented here are several data sets that gather information collected from the labels of the FDA approved drugs: their molecular structures and those of the described active metabolites, their associated pharmacokinetics and pharmacodynamics data, and the history of their marketing authorization by the FDA. To date, 1852 chemical structures have been identified with a molecular weight less than 2000 of which 492 are or have active metabolites. To promote the sharing of data, the original web server was upgraded for browsing the database and downloading the data sets (http://chemoinfo.ipmc.cnrs.fr/edrug3d). It is believed that the multidimensional chemistry-oriented collections are an essential resource for a thorough analysis of the current drug chemical space. The data sets are envisioned as being used in a wide range of endeavors that include drug repurposing, drug design, privileged structures analyses, structure−activity relationship studies, and improving of absorption, distribution, metabolism, and elimination predictive models.



KEYWORDS: *FDA-approved drugs, active metabolites, pharmacokinetics, pharmacodynamics, data sets, structure−activity relationships, cheminformatics*

Approved drugs are the most widely studied small molecules for their function and effects on humans as well as for what the body's physiology does to the molecules. They are a rich source of information to get insight into which properties are required for a molecule to be an administered drug and to optimize the search for new therapeutics. Drugs are characterized by two main properties: their pharmacodynamics (PD) and their pharmacokinetics (PK). The pharmacodynamics summarizes the mechanism of action, the biological targets, and their binding affinities. The pharmacokinetics informs on the fate of the therapeutic agent in the body and is characterized by its absorption, distribution, metabolism, and elimination (ADME). ADME properties represent whole body pharmacokinetics that integrates the numerous events of interaction between the drug and the organism molecular components such as proteins and membranes. ADME parameters are represented by the bioavailability (F) for absorption, the volume of distribution (VD) and the plasma protein binding (PPB) for distribution, and the clearance (Cl) and the half-life ($t_{1/2}$) for metabolism and excretion, each derived from the measurement of drug concentrations in blood or plasma. These experimental properties, although they may vary broadly between individuals, translate the drug interactions with the organism, which depend on the physicochemical properties of the drug such as the ionization state or the lipophilicity, to name just a few.[1] Pharmacodynamic and pharmacokinetic parameters are optimized during the drug discovery process relying on medicinal chemist experience and sometimes with the assistance of predictive models.[2−4] To foster the development of improved methods, a database and four manually curated data sets built from the Drugs@FDA data files are proposed.[5] e-Drug3D database is updated once a year using a semiautomatic method to update information on registered drugs and a manual processing for newly approved molecules.[6] A classification of drug names by the year of first approval is proposed on the home page; it aims to follow the evolution trend of approved small molecules over the years. For example, 31 drug structures were approved in 2015 and only 13 in 2016. The list of FDA products that are not registered in the database (biologics, contrast agents, electrolytes, etc.) are also made available.[7]

Over the years, several drug information databases have been designed: DrugBank,[8] ChEMBL/Drugs,[9] SuperDrug,[10] IDAAPM,[11] or DrugCentral.[12] These databases are useful resources on approved and experimental drugs for various reasons, which include drug and off-label indications, contraindications, dosage, bioactivity profiles, drug metabolism, to name just a few. However, except for the DrugBank, these databases do not report the experimental pharmacokinetic parameters, and none have explored the differential characterization of the structure of approved molecules and their active metabolites (Table S1). The active metabolite of the administrated drug is sometimes the only one to bind the primary target, and in some cases, both can possess biological activity. This information is required for those who want to establish meaningful structure−activity relationships. In addi-

Received: November 9, 2017
Accepted: January 29, 2018
Published: January 29, 2018

DOI: 10.1021/acsmedchemlett.7b00462
ACS Med. Chem. Lett. 2018, 9, 204−209

tion, the experimental pharmacokinetic parameters may be that of the drug and/or the active metabolite. For example, the bioavailability value is that of the administered drug structure, whereas the volume of distribution may be that of the active metabolite. This distinction cannot be performed by automatic processing of the drug label files because no such formalism exists. A case-by-case manual search is therefore needed. Of another significance, it is important to know which molecular structure will be tested in a biological assay, for example, in the case of a drug repurposing campaign, and, if needed, how the active metabolite can be generated.[13] In this database, the structures of active metabolites are therefore stored with their associated experimental properties as described in the drug label. Each molecular structure was manually cross-checked twice with the SciFinder/CAS database. In the present database, links were created between "mother" and "daughter" molecules. This lineage makes it possible to display the group of structures during browsing as well as to include the information in the derived data sets. One hundred and eighty-six such groups are listed online.[14] The analysis of this data set shows that 26% of drug structures are or have an active metabolite.

Pharmacokinetic Data Set. At present, there are few published pharmacokinetic data sets: PK/DB,[15] PKKB,[16] and the review by Obach et al.[17] PK/DB contains pharmacokinetic information for 1389 small compounds incorporating structurally diverse drug-like and lead-like molecules extracted from the literature. The database is available online for browsing, but authors do not offer downloadable structure−property data sets. PKKB provides pharmacokinetic information for 1685 drug and drug-like molecules and offers structure−property data sets containing Caco-2, LogBB, P-gp inhibitory, human intestinal absorption, and oral bioavailability values that may be complementary to this pharmacokinetic data set. At the time, the study by Obach et al. provided the largest publicly available data set of human pharmacokinetic data (VD, Cl, fu, MRT, and $t_{1/2}$) for 670 drug compounds.[17] The complete list of molecules with pharmacokinetic data, full references, and comments by the authors are available as supplemental data, but molecular structures are not provided. Nevertheless, in this work, the extracted pharmacokinetic parameters were systematically crosschecked with theirs. Four hundred and twenty-two drugs were common to both data sets.

To offer a large and high-quality collection, several experimental pharmacokinetic properties were manually extracted from drug labels, including the volume of distribution (VD), clearance (Cl), plasma protein binding percentage (PPB), terminal half-life ($t_{1/2}$), bioavailability (F), maximal concentration of drug in the blood (C_{max}) and time to reach the C_{max} (T_{max}), route(s) of administration, and the comment on the experimental solubility. For cases where there was no access to the label file or if the information was missing, these parameters were searched for in the review by Obach et al.,[17] in the French drug registry Vidal,[18] or in the literature. The number of experimental values for each parameter is listed in Table 1. Figure 1 presents univariate statistics for four experimental pharmacokinetic properties.

Volume of Distribution. VD is a nonphysiological term that represents the apparent volume into which a drug is distributed based on the concentrations measured in the blood. The approximate physiological volumes of body fluids are 40 L in an adult (25 L intracellular fluids, 15 L extracellular (12 L interstitial + 3 L plasma)). In the present data set, VD values

Table 1. Property Titles and Number of Experimental Values

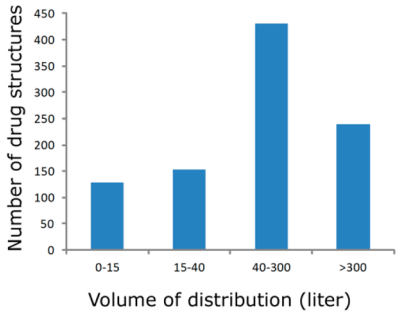
title	property	number
ID	e-Drug3D ID	1852
name	INN name	1852
CASRN	Chemical Abstract number	1850
year	year of approval (if identified)	1578[a]
status	if "discontinued"	382[b]
Pharmacokinetic Data Set		
is_a_metabolite	contains mother's ID	241[c]
has_a_metabolite	contains daughter's ID	264[c]
route	routes of administration	1693[d]
VD	volume of distribution (liter)	951
Cl	clearance (liter/hour)	939
$t_{1/2}$	half-life (hour)	1276
PPB	plasma protein binding (%)	1061
F	bioavailability (%)	524
C_{max}	maximal concentration in blood	766
C_{max_unit}	C_{max} unit	766
T_{max}	time to reach the C_{max} (h)	745
solubility	comments on solubility	957
Pharmacodynamic Data Set		
primary target	name of the primary target	1653
unified set of targets	ChEMBL, DrugCentral, and canSAR molecular drug targets[35]	1732
ATC code(s)	WHO ATC classification codes	1626
Registration Data Set		
natural compound	KnapSack ID, if it exists	78
FDA approvals	NDA numbers, commercial names, and companies	1693
class	general information on the therapeutic indication	1852

[a]Year = 0 when there is no identified year of approval or if it is an active metabolite without a NDA number (FDA approval number). [b]Discontinued structures are no longer marketed but still possess a NDA number. [c]Thirteen drugs are in both categories. [d]The 159 absent structures are active metabolites without NDA number.

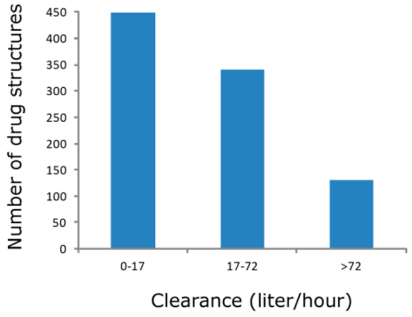
are expressed as liters. At the upper extreme limit, the drug Raloxifen has a VD value of 152 620 L, indicating a low blood concentration and an extensive level of tissue partitioning. At the lower extreme, Succinylcholine has a value of 0.14 L, indicating that the drug is concentrated in the blood and does not diffuse well in the whole organism. The mean and median values of this data set are 914 and 93 L, respectively. The majority of drugs (75%) have a VD value less than 300 L, and 30% have a value less than 40 L. VD is one of the fundamental PK parameters of drug candidates since it, together with clearance, determines the half-life that affects the dosing regimen of the drug.[19] The present experimental collection is expected to be of great use to assess and improve predictive models for VD.[4,20,21]

Clearance. The clearance Cl is the volume of blood that is cleared of the drug per unit of time. The approximate blood flow through the liver for adults is 72 L/h. Cl values range from 0.014 (Cytarabine, an antineoplastic) to 6900 (Alprostadil, prostaglandin E1) L/h. The mean and median values for this data set are 67 and 17 L/h, respectively. The majority of drugs (75%) have Cl values of less than 46 L/h, and 86% have values of less than 72 L/h. Clearance is an important pharmacokinetic parameter that is still difficult to model because it involves the prediction of the compound metabolic stability.[4,22] Nevertheless, computational approaches can be successful in predicting the most probable metabolized sites of structures.[3]

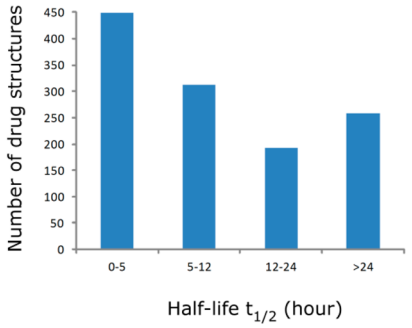
DOI: 10.1021/acsmedchemlett.7b00462
ACS Med. Chem. Lett. 2018, 9, 204−209



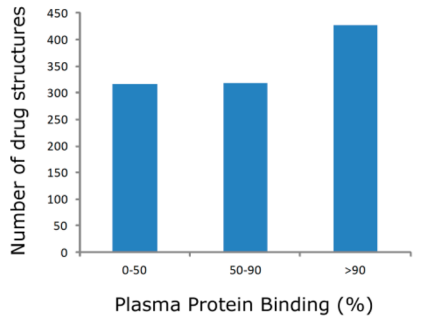
a) N = 951; Range = [0.14 - 152620]; Mean = 914; Median = 93
25th percentile = 30; 75th percentile = 304; 95th percentile = 2001



b) N = 939; Range = [0.014 - 6900]; Mean = 67; Median = 17
25th percentile = 6; 75th percentile = 46; 95th percentile = 181



c) N = 1276; Range = [0.03 - 87600]; Mean = 107; Median = 7
25th percentile = 2.50; 75th percentile = 20; 95th percentile = 102



d) N = 1061; Range = [0.0 - 99.90]; Mean = 68; Median = 83
25th percentile = 42; 75th percentile = 97; 95th percentile = 99

Figure 1. Univariate statistics for experimental PK parameters: (a) VD, (b) Cl, (c) $t_{1/2}$, and (d) PPB. The number of values, the range, the mean, the medium, the first quartile, the third quartile, and the 95th percentile values are indicated under the graph.

The modification of these metabolic liable sites may improve the intrinsic clearance of a molecule together with the VD and half-life as shown in the example of the drugs Betaxolol and Metoprolol.[19,23]

Half-Life. The half-life ($t_{1/2}$) is the period of time required for the concentration or amount of drug in the body to be reduced by one-half. After the passing of 7 half-times, the drug is 99% eliminated from the body, assuming a single dose. Therefore, a $t_{1/2}$ value lower than 5 means that there is an almost complete clearance over 24 h. In the present data set, 40% of drugs possess a $t_{1/2}$ value lower than 5 h. Half-life values range from 0.03 (~2 min) for Mivacurium to 87 600 h (~10 years) for Alendronate because this drug is sequestrated into bone. The mean and median values for this data set are 107 and 7 h, respectively. This property is not usually predicted by quantitative structure−activity methods, but it is part of physiologically based pharmacokinetic modeling (PBPK) approaches.[4,24,25] PBPK models incorporate *in vitro* physicochemical and biochemical data in a physiologically based approach to predict plasma concentration−time curves.

Plasma Protein Binding. Recently, Liu et al. argued that PPB should not be optimized during drug design because theoretical analyses and experimental observations show that low plasma protein binding does not necessarily lead to high *in vivo* unbound plasma concentration.[26] In line with their observations, the analysis presented here of the distribution of PPB values shows that 50% of drug structures have a value of PPB greater than 83%, and 32% have a value of PPB greater than 95%.

Maximal Concentration in Blood. The maximal concentration in blood values (C_{max}) spans a wide range varying from 711 pM for the oral antiparkinsonian drug Bromocriptine to millimolar concentration for the antihyperammonemic oral drug Phenylbutyrate and the oral/injected antifibrinolytic Aminocaproic acid. For the 754 drugs in the millimolar and nanomolar concentration range, the median and average C_{max} values are 513 and 35 μM, respectively. These concentration values are comparable to that of alpha-1 acid glycoprotein in plasma (17.5 μM) and human serum albumin (600 μM).[27,28]

In this data set, 417 drug structures fall in the micromolar range with a mean C_{max} value of 64 μM, and 337 compounds fall in the nanomolar range with a mean C_{max} value of 179 nM.

C_{max} is not a property intended for quantitative structure−activity modeling. Its interest lies in the evaluation of the maximum concentration reached at time T_{max} by the therapeutic agent. The common value of T_{max} is approximately 2 h. C_{max} values give an indication of the possibility of interaction with other proteins and/or off-targets whose affinity is of the same order of magnitude.[29−31]

Solubility. Through extensive mining of drug labels, comments on the experimental solubility were found for 957 drugs structures. This analysis allowed the categorization of the data into two groups: those that are characterized as "insoluble, not soluble, slightly, sparingly, or poorly soluble", or with a solubility value less than <1 mg/mL in water, and the others. The first group is represented by 56% of compounds (540 drugs) and the second by 44% of compounds (417 drugs). In contrast to common belief, many approved drugs are not soluble in water[32] despite the fact that more than 75% of compounds in both groups possess an oral route of administration. Many predictive solubility methods have been developed, but the results are far from satisfactory,[3,33] and other alternatives to optimize the solubility have been proposed.[34]

Complementary Data Sets. In addition to the structures and pharmacokinetic parameters, information on their primary target and the history of their marketing was assembled.

DOI: 10.1021/acsmedchemlett.7b00462
ACS Med. Chem. Lett. 2018, 9, 204−209

Pharmacodynamic Data Set. The name of the primary drug target and/or the mechanism of action were manually extracted from drug labels or mined from the literature. Additional target annotations were extracted from the recently released unified set of drug efficacy targets.[35] These assignments include links to the ChEMBL and UniProt databases to facilitate drug-target analyses like, for example, docking studies. Next, the WHO Anatomical Therapeutic Chemical (ATC) codes were automatically retrieved by matching the drug names.[36] ATC codes were found for 1626 drug structures. As expected, some active metabolites, old molecules, and newly approved drugs were not indexed. ATC codes allow the association of drugs with their primary target, therapeutic class(es), and indications. The pharmacodynamic information is of value when studying specific groups of drugs (antibiotics, antihypertensives, etc.) or protein targets (G-protein-coupled receptors (GPCRs), nuclear receptors, etc.). Figure 2 shows the distribution of drug structures into the ten most commonly drugged protein families. This distribution is consistent with previous studies.[37,38] Enzymes constitute the largest group of primary targets containing 32% of all molecules. GPCRs and ion channels are the second (24% of drugs) and third (12%) largest target classes, respectively.

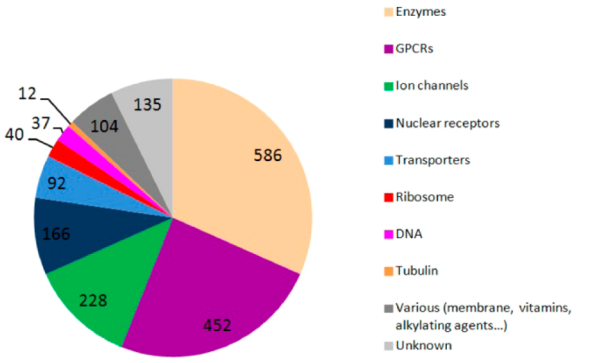


Figure 2. Primary target family distribution of the 1852 drug structures.

Registration Data Set. As noted above, information from the Drugs@FDA data files for each registered drug structure were automatically collected.[5] These files contain general information on approved drugs: INN name, application number (also called NDA number), submission date, commercial names, sponsor name, marketing status, routes of administration, and dosage. These files allowed extraction of years of approvals and tracing the history of the marketing authorization by the FDA.

As indicated by the associated references, one of the applications is the possibility to analyze the evolution over the years of the number of approved drugs,[39,40] the physicochemical properties,[41] the structural diversity,[42] the privileged scaffolds,[43] the occurrence of natural products,[44,45] the targeted protein families,[37,38] the occupancy of the therapeutic areas,[46−48] or the organizations obtaining approvals for new drugs or generics.[40]

Drug Labels. A label contains an accurate summary of the essential scientific information needed for the safe and effective use of a FDA-approved drug. As previously indicated, the pharmacokinetic and pharmacodynamic parameters from these files were extracted. Each approval has a unique number (NDA) that is used to name the label file. Label files are in PDF format and are usually available online at the U.S. Food and Drug Administration Web site (see ref 49), where the last character 'X' must be replaced by the NDA number.

In order to facilitate the analyses, the different label files for each drug structure in the registration data set were automatically collected and stored and a compressed directory available for download was generated. Of note, some drug labels, especially for old molecules, may no longer exist at the FDA Web site, but they may be viewed at Dailymed[50] or Drugs.com.[51] Therefore, one may easily access the label file of the drug from which experimental information was extracted. In further studies, it would be interesting to use and assess techniques that automatically extract formalized information from PDF files to search for additional properties such as adverse drug reactions or drug metabolism enzymes.[52]

In summary, several structure−activity/property data sets were created and aimed to perform retrospective analyses of past successes and to foster the development of improved predictive methods relevant to the drug discovery and optimization process. In particular, to the best knowledge, the pharmacokinetic data set containing chemical structures, experimental parameters and active metabolite lineage represents the largest publicly available collection tailored to meet the requirements of structure−activity relationship studies. It is believed that these chemistry-oriented collections should provide valuable resources with a sufficient number of high quality experimental data for further analyses and developments in the field.

EXPERIMENTAL PROCEDURES

e-Drug3D Database and Web Server. The implementation and organization of the database as well as the web interface were described previously.[6] At present, the updated database contains eight additional pharmacokinetic parameters (comment on solubility, VD, Cl, $t_{1/2}$, PPB, F, C_{max}, and T_{max}) and an additional pharmacodynamic annotation extracted from the ChEMBL, DrugCentral, and canSAR joint effort.[35] The web server further evolved with a new structure editor (JSME[53]) and with new interactive visualization editors (ChemDoodle[54] and JSmol[55]) to replace Java applications. e-Drug3D gives access to the structures of drugs and active metabolites approved from 1939 to August 2017. Each molecular structure possesses a unique identifier (ID) that identifies its chemical structure and properties across data sets. The database is updated every year, and it currently contains 1453 unique INN names that are represented by 1852 different structures because active metabolites and enantiomers have their own entry in the database. All enantiomers possess the same INN name.

Structure Data Set. The structure data set was generated as previously described.[6] It contains 1852 molecular structures with 3D coordinates in SDF format. Each molecular structure was manually cross-checked twice with the SciFinder/CAS database and with the molecular structure depicted in the drug label. The datablock of the SDF file contains the INN name of the drug, the CAS registry number (CASRN), the ID, and the status, if the drug is discontinued. When the INN name of the active metabolite did not exist, its name was set by adding the prefix 'M' to the name of the associated marketed drug such as, for example, SELEXIPAG and M SELEXIPAG, its active metabolite. All data sets are available to download from http://chemoinfo.ipmc.cnrs.fr/MOLDB/.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsmedchemlett.7b00462.

DOI: 10.1021/acsmedchemlett.7b00462
ACS Med. Chem. Lett. 2018, 9, 204−209

Table S1 comparing the different databases and data sets (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: douguet@ipmc.cnrs.fr.

ORCID

Dominique Douguet: 0000-0001-6209-6464

Funding

This work was supported by the Centre National de la Recherche Scientifique (CNRS) and the Institut National de la Santé et de la Recherche Médicale (Inserm).

Notes

The author declares no competing financial interest.

ACKNOWLEDGMENTS

I am grateful to Dr. Bruno Villoutreix and Dr. Amanda Patel for helpful discussions and for proofreading and correction of this manuscript. I thank D. Barbier for technical assistance with the web server.

ABBREVIATIONS

FDA, U.S. Food and Drug Administration; INN, international nonproprietary name; PK, pharmacokinetics; PD, pharmacodynamics; ADME, absorption, distribution, metabolism, and excretion; VD, volume of distribution; Cl, clearance; F, bioavailability; PPB, plasma protein binding; fu, fraction of unbound drug in plasma; MRT, mean residence time; $t_{1/2}$, half-life; C_{max}, maximal concentration of drug in the blood; T_{max}, time to reach the C_{max}; logBB, brain−blood permeability; P-gp, P-glycoprotein; PBPK, physiologically based pharmacokinetic modeling; WHO, World Health Organization; ATC, anatomical therapeutic chemical; CASRN, Chemical Abstract registry number; NDA, new drug application.

REFERENCES

(1) Gleeson, M. P. Generation of a set of simple, interpretable ADMET rules of thumb. *J. Med. Chem.* **2008**, *51* (4), 817−834.

(2) Jonsdottir, S. O.; Jorgensen, F. S.; Brunak, S. Prediction methods and databases within chemoinformatics: emphasis on drugs and drug candidates. *Bioinformatics* **2005**, *21* (10), 2145−2160.

(3) Wang, Y.; Xing, J.; Xu, Y.; Zhou, N.; Peng, J.; Xiong, Z.; Liu, X.; Luo, X.; Luo, C.; Chen, K.; Zheng, M.; Jiang, H. In silico ADME/T modelling for rational drug design. *Q. Rev. Biophys.* **2015**, *48* (4), 488−515.

(4) van de Waterbeemd, H.; Gifford, E. ADMET in silico modelling: towards prediction paradise? *Nat. Rev. Drug Discovery* **2003**, *2* (3), 192−204.

(5) Drugs@FDA data files. https://www.fda.gov/Drugs/InformationOnDrugs/ucm079750.htm.

(6) Pihan, E.; Colliandre, L.; Guichou, J. F.; Douguet, D. e-Drug3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design. *Bioinformatics* **2012**, *28* (11), 1540−1541.

(7) excluded_NDA.out. http://chemoinfo.ipmc.cnrs.fr/MOLDB/excluded_NDA.out.

(8) Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A. C.; Wishart, D. S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **2011**, *39* (Database), D1035−D1041.

(9) Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrian-Uhalte, E.; Davies, M.; Dedman, N.; Karlsson, A.; Magarinos, M. P.; Overington, J. P.; Papadatos, G.; Smit, I.; Leach, A. R. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45* (D1), D945−D954.

(10) Goede, A.; Dunkel, M.; Mester, N.; Frommel, C.; Preissner, R. SuperDrug: a conformational drug database. *Bioinformatics* **2005**, *21* (9), 1751−1753.

(11) Legehar, A.; Xhaard, H.; Ghemtio, L. IDAAPM: integrated database of ADMET and adverse effects of predictive modeling based on FDA approved drug data. *J. Cheminf.* **2016**, *8*, 33.

(12) Ursu, O.; Holmes, J.; Knockel, J.; Bologa, C. G.; Yang, J. J.; Mathias, S. L.; Nelson, S. J.; Oprea, T. I. DrugCentral: online drug compendium. *Nucleic Acids Res.* **2017**, *45* (D1), D932−D939.

(13) Baker, M. Reproducibility: Check your chemistry. *Nature* **2017**, *548* (7668), 485−488.

(14) metabolite_groups. http://chemoinfo.ipmc.cnrs.fr/MOLDB/metabolite_groups.html.

(15) Moda, T. L.; Torres, L. G.; Carrara, A. E.; Andricopulo, A. D. PK/DB: database for pharmacokinetic properties and predictive in silico ADME models. *Bioinformatics* **2008**, *24* (19), 2270−2271.

(16) Cao, D.; Wang, J.; Zhou, R.; Li, Y.; Yu, H.; Hou, T. ADMET evaluation in drug discovery. 11. PharmacoKinetics Knowledge Base (PKKB): a comprehensive database of pharmacokinetic and toxic properties for drugs. *J. Chem. Inf. Model.* **2012**, *52* (5), 1132−1137.

(17) Obach, R. S.; Lombardo, F.; Waters, N. J. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. *Drug Metab. Dispos.* **2008**, *36* (7), 1385−1405.

(18) Vidal. https://www.vidal.fr.

(19) Smith, D. A.; Beaumont, K.; Maurer, T. S.; Di, L. Volume of Distribution in Drug Design. *J. Med. Chem.* **2015**, *58* (15), 5691−5698.

(20) Lombardo, F.; Obach, R. S.; Shalaeva, M. Y.; Gao, F. Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data. *J. Med. Chem.* **2002**, *45* (13), 2867−2876.

(21) Gleeson, M. P.; Waters, N. J.; Paine, S. W.; Davis, A. M. In silico human and rat Vss quantitative structure-activity relationship models. *J. Med. Chem.* **2006**, *49* (6), 1953−1963.

(22) Schneider, G.; Coassolo, P.; Lave, T. Combining in vitro and in vivo pharmacokinetic data for prediction of hepatic drug clearance in humans by artificial neural networks and multivariate statistical techniques. *J. Med. Chem.* **1999**, *42* (25), 5072−5076.

(23) Manoury, P. M.; Binet, J. L.; Rousseau, J.; Lefevre-Borg, F.; Cavero, I. G. Synthesis of a series of compounds related to betaxolol, a new beta 1-adrenoceptor antagonist with a pharmacological and pharmacokinetic profile optimized for the treatment of chronic cardiovascular diseases. *J. Med. Chem.* **1987**, *30* (6), 1003−1011.

(24) Huang, W.; Lee, S. L.; Yu, L. X. Mechanistic approaches to predicting oral drug absorption. *AAPS J.* **2009**, *11* (2), 217−224.

(25) Sager, J. E.; Yu, J.; Ragueneau-Majlessi, I.; Isoherranen, N. Physiologically Based Pharmacokinetic (PBPK) Modeling and Simulation Approaches: A Systematic Review of Published Models, Applications, and Model Verification. *Drug Metab. Dispos.* **2015**, *43* (11), 1823−1837.

(26) Liu, X.; Wright, M.; Hop, C. E. Rational use of plasma protein and tissue binding data in drug design. *J. Med. Chem.* **2014**, *57* (20), 8238−8248.

(27) Blain, P. G.; Mucklow, J. C.; Rawlins, M. D.; Roberts, D. F.; Routledge, P. A.; Shand, D. G. Determinants of plasma alpha 1-acid glycoprotein (AAG) concentrations in health. *Br. J. Clin. Pharmacol.* **1985**, *20* (5), 500−502.

(28) Ghuman, J.; Zunszain, P. A.; Petitpas, I.; Bhattacharya, A. A.; Otagiri, M.; Curry, S. Structural basis of the drug-binding specificity of human serum albumin. *J. Mol. Biol.* **2005**, *353* (1), 38−52.

(29) Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijer, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. Predicting new molecular targets for known drugs. *Nature* **2009**, *462* (7270), 175−181.

(30) Campillos, M.; Kuhn, M.; Gavin, A. C.; Jensen, L. J.; Bork, P. Drug target identification using side-effect similarity. *Science* **2008**, *321* (5886), 263−266.

DOI: 10.1021/acsmedchemlett.7b00462
ACS Med. Chem. Lett. 2018, 9, 204−209

(31) Hodos, R. A.; Kidd, B. A.; Shameer, K.; Readhead, B. P.; Dudley, J. T. In silico methods for drug repurposing and pharmacology. *Wiley Interdiscip Rev. Syst. Biol. Med.* **2016**, *8* (3), 186−210.
(32) Baell, J. B. Screening-based translation of public research encounters painful problems. *ACS Med. Chem. Lett.* **2015**, *6* (3), 229−234.
(33) Wang, J.; Hou, T. Recent advances on aqueous solubility prediction. *Comb. Chem. High Throughput Screening* **2011**, *14* (5), 328−338.
(34) Ishikawa, M.; Hashimoto, Y. Improvement in aqueous solubility in small molecule drug discovery programs by disruption of molecular planarity and symmetry. *J. Med. Chem.* **2011**, *54* (6), 1539−1554.
(35) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; Overington, J. P. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discovery* **2017**, *16* (1), 19−34.
(36) WHO Anatomical Therapeutic Chemical (ATC). https://www.whocc.no/atc_ddd_index.
(37) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discovery* **2006**, *5* (12), 993−996.
(38) Kinch, M. S.; Hoyer, D.; Patridge, E.; Plummer, M. Target selection for FDA-approved medicines. *Drug Discovery Today* **2015**, *20* (7), 784−789.
(39) Munos, B. Lessons from 60 years of pharmaceutical innovation. *Nat. Rev. Drug Discovery* **2009**, *8* (12), 959−968.
(40) Kinch, M. S.; Haynesworth, A.; Kinch, S. L.; Hoyer, D. An overview of FDA-approved new molecular entities: 1827−2013. *Drug Discovery Today* **2014**, *19* (8), 1033−1039.
(41) Proudfoot, J. R. The evolution of synthetic oral drug properties. *Bioorg. Med. Chem. Lett.* **2005**, *15* (4), 1087−1090.
(42) Vitaku, E.; Smith, D. T.; Njardarson, J. T. Analysis of the structural diversity, substitution patterns, and frequency of nitrogen heterocycles among U.S. FDA approved pharmaceuticals. *J. Med. Chem.* **2014**, *57* (24), 10257−10274.
(43) Welsch, M. E.; Snyder, S. A.; Stockwell, B. R. Privileged scaffolds for library design and drug discovery. *Curr. Opin. Chem. Biol.* **2010**, *14* (3), 347−361.
(44) Camp, D.; Garavelas, A.; Campitelli, M. Analysis of Physicochemical Properties for Drugs of Natural Origin. *J. Nat. Prod.* **2015**, *78* (6), 1370−1382.
(45) Patridge, E.; Gareiss, P.; Kinch, M. S.; Hoyer, D. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discovery Today* **2016**, *21* (2), 204−207.
(46) Kinch, M. S.; Patridge, E. An analysis of FDA-approved drugs for infectious disease: HIV/AIDS drugs. *Drug Discovery Today* **2014**, *19* (10), 1510−1513.
(47) Kinch, M. S.; Patridge, E.; Plummer, M.; Hoyer, D. An analysis of FDA-approved drugs for infectious disease: antibacterial agents. *Drug Discovery Today* **2014**, *19* (9), 1283−1287.
(48) Kinch, M. S. An analysis of FDA-approved drugs for oncology. *Drug Discovery Today* **2014**, *19* (12), 1831−1835.
(49) Drugs@FDA: FDA Approved Drug Products. https://www.accessdata.fda.gov/Scripts/cder/daf/index.cfm?event=overview.process&ApplNo=X.
(50) Dailymed. https://dailymed.nlm.nih.gov.
(51) Drugs.com. https://www.drugs.com.
(52) Swain, M. C.; Cole, J. M. ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature. *J. Chem. Inf. Model.* **2016**, *56* (10), 1894−1904.
(53) Bienfait, B.; Ertl, P. JSME: a free molecule editor in JavaScript. *J. Cheminf.* **2013**, *5*, 24.
(54) Burger, M. C. ChemDoodle Web Components: HTML5 toolkit for chemical graphics, interfaces, and informatics. *J. Cheminf.* **2015**, *7*, 35.
(55) Hanson, R. M.; Prilusky, J.; Renjian, Z.; Nakane, T.; Sussman, J. L. JSmol and the next-generation web-based representation of 3D molecular structure as applied to Proteopedia. *Isr. J. Chem.* **2013**, *53* (3−4), 207−216.

DOI: 10.1021/acsmedchemlett.7b00462
ACS Med. Chem. Lett. 2018, 9, 204−209